KEMENTERIAN SAINS,
TEKNOLOGI DAN INOVASI
MINISTRY OF SCIENCE, TECHNOLOGY AND INNOVATION

GUIDELINES ON
OPEN SCIENCE (OS) IN PUBLIC FUNDED RESEARCH

Malaysia Open Science Platform (MOSP)

GUIDELINES ON

OPEN SCIENCE (OS) IN PUBLIC FUNDED RESEARCH

GUIDELINES ON OPEN SCIENCE (OS) IN PUBLIC FUNDED RESEARCH
These Guidelines were prepared as a deliverable for the Malaysia Open Science Alliance Working Group on Guidelines.

Academy of Sciences Malaysia
Level 20, West Wing, MATRADE Tower
Jalan Sultan Haji Ahmad Shah off Jalan Tuanku Abdul Halim
50480 Kuala Lumpur, Malaysia

Perpustakaan Negara Malaysia Cataloguing-In-Publication Data
Guidelines on Open Science (OS) in Public Funded Research

ISBN 978-983-2915-63-8
Keywords related to publication
provided by National Library during
ISBN application process
1.Information storage and retrieval system–Science
2.Information storage and retrieval system–Science and state.
3.Science and state– Malaysia
4.Research–Malaysia
5. Government publications–Malaysia
I. Akademi Sains Malaysia

TABLE OF CONTENT

FOREWORD

The Malaysia Open Science Alliance (MOSA) in collaboration with the Academy of Sciences Malaysia (ASM) has been tasked by the Government in drafting a Guidelines on Open Science in Public Funded Research. The mandate given was in line with the Open Science Initiative championed by MOSP. The proposal for the guidelines was drafted in collaboration and consultation with various stakeholders through a series of information and dialogue meetings. The views that emerged during these meetings were taken into account in the work on producing the proposal for guidelines

The guidelines will help to ensure **scholarly publications and research data** resulting from **publicly funded research** to be openly and publicly available. Information and Communication Technologies (ICTs) as an enabler, has however made it easier and doable. As such openness in obtaining, processing, publishing and disseminating research information has become easily achievable due to the spread of ICTs and ICT-enabled services.

In addition, there are some socio-economic benefits and diverse opportunities to be derived from Open Science. Perhaps the most important reasons are the broad economic benefits and growth, both public and private. Scholarly publications and research data made available and accessible through Open Science have been shown to be economic force enhancers and multipliers, creating value many times over and providing much greater returns on public research investments. The generative or pro-creative effects as a result of Open Science are key in this regard.

Undoubtedly, Open Science will have an effect on society's social welfare. Not only will it meet society's expectations on appropriate management of Open Science assets and resources, it will also provide diverse reputational gains apart from incorporating ethical principles for accessing and using scholarly publications and research data. In the public research it ca substantially reduce unproductive barriers to interdisciplinary, inter-institutional, and international research. Besides n enabling data mining for knowledge discovery in a growing sea of big data, Open Science is essential for the verification of research results and in generating broad trust in them. It avoids many inefficiencies, such as the unnecessary duplication of research and the identification of erroneous results. Open Science will promote more research and new types of research. It also permits the legal interoperability of data when multiple sources of data are combined for new knowledge.

Finally, Open Science will help to improve governance. Public data made openly available through the public institutional portals will support improved decision-making and transparency in government and society. For a developing economy like Malaysia, Open Science will help to build freedom in society, and trust in governance and its many functions.

ACKNOWLEDGEMENT

We would like to extend our appreciation and thanks to all those that have participated in the study. We acknowledged with thanks the comments and suggestions imparted and guidance extended during the series of engagement with the Malaysia Open Science Alliance and its Working Group on Guidelines, the Academy of Sciences Management/ Secretariat, the management, the researchers and the librarians of the selected Malaysian public and private institutions of higher learning (IHL) and the public research institutions (PRIs).

MALAYSIA OPEN SCIENCE ALLIANCE, WORKING GROUP ON POLICY AND SECRETARIAT/ASM MANAGEMENT

MALAYSIA OPEN SCIENCE ALLIANCE

1. Professor Dr Noorsaadah Abd Rahman FASc (Chairperson, UM)
2. Professor Datuk Dr Rohana Yusof FASc (ASM)
3. Academician Professor Emerita Datuk Dr Mazlan Othman FASc (ISC-ROAP)
4. Professor Dr Shaliza Ibrahim (UM)
5. Professor Dr Ismanizan Ismail (UKM)
6. Professor Dr Zulkifli Idrus (UPM)
7. Professor Dr Rosni Abdullah (USM)
8. Professor Dr Abrizah Abdullah (UM)
9. Puan Roshana Alma Mohd Ali (MOSTI)
10. Puan Noraziah Sharuddin (UTM)
11. Dr Liew Chee Sun (UM)
12. Dr Salina Abdul Aziz (MOH)
13. Encik Azih Yusof (MAMPU)
14. Puan Zaiha Mat Nor (MAMPU)
15. Encik Kamaruhzaman Mat Zin (MASTIC)

MALAYSIA OPEN SCIENCE ALLIANCE WORKING GROUP ON GUIDELINES

1. Professor Datuk Dr Rohana Yusof FASc (Chairperson, ASM)
2. Emeritus Professor Dato' Dr Rahmah Mohamed FASc (ASM)
3. Dr Khasiah Zakaria (UM)
4. Dr Sik Cheng Peng (UM)
5. Dr Nasuha Lee Abdullah (USM)
6. Dr Syed Norris Hikmi Syed Abdullah (UTM)
7. Puan Roshana Alma Ali (MOSTI)
8. Encik Kamaruhzaman Mat Zin (MASTIC)
9. Encik Kamarulzaman Kamaruddin (MOSTI)
10. Professor Dr Ismanizan Ismail (UKM)
11. Professor Dr Samsilah Roslan (UPM)

ACADEMY OF SCIENCES MALAYSIA - MALAYSIA OPEN SCIENCE PLATFORM

1. Puan Hazami Habib (Chief Executive Officer)
2. Dr Nurzatil Sharleeza Mat Jalaluddin (Senior Executive)
3. Nurlizah Abu Bakar (Senior Executive)
4. Shyama Sundari Devi (MOSP Executive)
5. Nur Hanisah Ismail (MOSP Executive)
6. Nur Faizah Azhari (MOSP Executive)
7. Shaanthana Subramaniam (MOSP Executive)
8. Muhamad Fariz Muhamad Suhaimi (MOSP Executive)

LIST OF FIGURES

LIST OF TABLES

LIST OF BOXES

ABBREVIATION AND ACRONYMS

APCs	Article Processing Charges
API	Application Programming Interface
ARDC	Australian Research Data Commons
BPCs	Book Processing Charges
DMP	Data Management Plan
DOAJ	Directory of Open Access Journals
DOI	Digital Object Identifier
DVD	Digital Versatile Disc
EOSC	European Open Science Cloud
FAIR	Findable, Accessible, Interoperable and Reusable
FOSTER	Facilitate Open Science Training for European Research
ICT	Information and Communications Technology
ID	Identification/Identity
IHLs	Institute of Higher Learnings
IP	Intellectual Property
ISC	International Science Council
ISO	International Organization for Standardization
KPI	Key Performance Indicator
MOHE	Ministry of Higher Education
MOSP	Malaysia Open Science Platform
MOSTI	Ministry of Science, Technology, and Innovation
NCBI	National Centre for Biotechnology Information
NGO	Non-Government Organisation
NPSTI	National Policy on Science, Technology and Innovation
OECD	Organisation for Economic Co-operation and Development
ORCID	Open Researcher and Contributor ID
PLOS	Public Library of Science
PIs	Principal Investigators
PRI	Public Research Institutes
RAND	RAND Corporation
RDM	Research Data Management
RI	Research Institutes
RU	Research University
SSH	Social Science and Humanities
STEM	Science, Technology, Engineering and Mathematics
STI	Science, Technology and Innovation
UKM	Universiti Kebangsaan Malaysia
UM	Universiti Malaya
UPM	Universiti Putra Malaysia
US	United States
USM	Universiti Sains Malaysia
UTM	Universiti Teknologi Malaysia
VRE	Virtual Research Environment

GLOSSARY

Article Processing Charges (APCs): fees that some scholarly publishers charge authors of academic papers to publish their work in open access.

Book Processing Charges (BPCs): fees charged by a publisher to make a book open access.

Confidential: highly restricted information due to the law such as Data Protection, policy, agreement or duty of confidence arising from the nature of relationship between the parties. Inappropriate disclosure of the information would be likely to cause serious damage or distress to individuals and/or constitute unfair/unlawful processing of "sensitive personal data" under the Data Protection Act; and/or seriously damage the government and institution interests and reputation; and/or significantly threaten national security.

Copyrights: collection of legal rights that are attached to an original work when it is created. Copyright allows the copyright owner to control certain acts to do with their work (e.g. copying) and to prevent others from using the protected material without permission.

Data Curator: responsible for organising and integrating data collected from various sources. It involves publication, presentation, reuse and preservation of the data.

Data Custodian: Data owners are also data custodians who own the data storage facilities. A data custodian is an IT individual or organisation responsible for the IT infrastructure providing and protecting data in conformance with the policies and practices prescribed by data governance.

Data Governance: A cross-functional management programme that treats data as an organisational asset through the collection of policies, standards, processes, people and technologies to achieve a set of goals.

Data Management Plan (DMP): a living document that records how the research data arising from the research project will be handled during and after the project is completed, describing what data will be shared and/or made open, and how it will be curated and preserved.

Data Originators: Researchers who produce research data and who are credited for their work. Also known as data creator.

Data Owner: Institutions, which also are employers of researchers, or the research institutions receiving and administering the grants.

Digital Repository: an on-line archive for collecting, preserving and disseminating digital copies of the intellectual research outputs.

Data Sharing: Data can be shared at any time either publicly or privately among collaborators, while the proper documentation and code is open source to ensure that others can build on and benefit from.

Data Steward: Protects the integrity and quality of data, adherence or compliance to standards and protocols, governance and advocacy. The role of data stewards complements curators in the aspects of both metadata management activities and data governance.

Data User: Individuals who re-use data and have responsibilities to acknowledge the sources of their data by citation or giving appropriate credits to data originators.

Errata: a list of errors and their corrections inserted, usually on a separate page or slip of paper, in a book or other publication. This is also referred to as corrigenda.

Embargo: the period during which a publication can be ‘closed’ while deposited in the repository (i.e. the publication is not openly available).

FAIR Data Principles: refers to a set of principles to make data Findable, Accessible, Interoperable and Reusable for scientific management, data stewardship and Open Science framework.

Gold Open Access: makes the final published version of an article freely available and permanently accessible for everyone, immediately after publication.

Green Open Access: also known as "self-archiving", it is "the practice of placing a version of an author's manuscript into a repository, making it freely accessible for everyone."[1] . The version (pre-print or post-print) that can be deposited into a repository is dependent on the funder or publisher.

Metadata: means "data about data". Metadata are the descriptors used for describing, tracing, use and management of the deposited item. Metadata describes characteristics such as content, quality, format, location and contact information.

Open Access: it's freely availability on the public internet, permitting any users to read, download, copy, distribute, print, search or link to the full texts of these articles, crawl them for indexing, pass them as data to software or use them for any other lawful purpose without financial, legal or technical barriers other than those inseparable from gaining access to the internet itself.

Open Data: is defined in essence, as data that can be freely used, re-used and redistributed by anyone. Besides being commonly associated with Open Government Data, Open Data also refers to Open Business Data and Citizen Generated Data. The main criteria for Open data are complete, primary, timely, accessible, machine-processable, non-discriminatory, non-proprietary and license-free.

Open Peer Review: a scholarly review mechanism where both the identities of the reviewer and the author are known to one another during the review and publication process.

Pre-print: refers to the version of an academic paper which is submitted by an author for peer review.

Post-print: refers to the final version of an academic paper before publication, incorporating the revisions made as a result of the peer review process or as accepted for publication if no changes were made.

Research: defined as any creative and systematically performed work with the goal of furthering knowledge.

Research data: any information that has been collected, observed, generated or created to validate original research findings. Although usually digital, research data also includes non-digital formats.

Research Data Lifecycle: consists of data acquisition, processing, analysis, curation, sharing and re-use. The data life cycle is divided into two domains i.e. private (green-colour coded) and public (blue-colour coded).

Research Data Management (RDM): concerning the organisation of data, from its entry to the research cycle through to the dissemination and archiving of valuable results. It aims to ensure reliable verification of results, and permits new and innovative research built on existing information.

Restricted data: data that is restricted or prohibited from disclosure. Restricted data would include confidential data. In some circumstances, access to sensitive data can be restricted, depending on whether there is any express prohibition or policy discouraging its disclosure.

Sensitive data: data that can be used to identify an individual, species, object, process, or location that introduces a risk of discrimination, harm, or unwanted attention. Under law and the research ethics governance of most institutions, sensitive data cannot typically be shared in this form, with few exceptions.

---

[1] https://www.springer.com/gp/authors-editors/authorandreviewertutorials/open-access/what-is-open-access/10286522

CHAPTER 1 INTRODUCTION

1.1. BACKGROUND

Innovative scientific research through public research funding has a crucial role in addressing national and global challenges. But this research is only meaningful if it is translated and depends a lot on how it is fostered. Fostering collaborative exchanges through Open Science between different scientific and research communities and other communities at large and assuring its widest dissemination in terms of speed and depth is also crucial too. In short, the exchange of ideas, knowledge and data emerging scientific research through Open Science is vital for country's progress and development in terms of knowledge creation, wealth creation and societal well-being. Thus, making scientific research and data, transparent and accessible to all is core.

Open Science is an initiative to make research output such as data and publications more transparent and accessible. It is about extending the principles of openness to the whole research cycle based on cooperative work and new ways of diffusing knowledge through digital technologies and collaborative tools1. For those research outputs to be accessible and can be shared by "everyone", they should be properly managed and curated, meeting the principles of Findable, Accessible, Interoperable and Reusable (FAIR). With FAIR data, researchers are able to create, share and re-use quality, valuable, high integrity and responsible data, fueling scientific progress to its fullest potential.

Admittedly, Open Science is gaining worldwide consensus as more countries have introduced and implemented the initiative at the national and regional levels. In Malaysia, Open Science is introduced through the Malaysia Open Science Platform (MOSP), an initiative managed by the Academy of Sciences Malaysia through the Malaysia Open Science Alliance, and funded by the Ministry of Science, Technology and Innovation (MOSTI). MOSP aims to gather and consolidate Malaysia's research data which are valuable national assets in a platform that would enable accessibility and sharing of these research data in accordance with the **FAIR** principle. In short, open data sharing is the way while embracing the FAIR principles. This Platform represented as a strategic transformative initiative to strengthen Malaysia's STI collaborative ecosystem towards achieving Shared Prosperity Vision 2030 and addressing the United Nations Sustainable Development Goals.

In this regard, MOSP has embarked on a two-year pilot project, launched on 7 November 2019. It has tasked the Malaysia Open Science Alliance is to look into three main areas, which are (1) Guidelines, (2) Capacity Building and Awareness and (3) Infrastructure. The development of a **Guidelines on Open Data Sharing in Research** is one such commitment which is in line with one of the strategies in the **Policy on Science, Technology and Innovation (NPSTI 2021-2030)**. The formulation of the Guidelines document has taken into account and consolidated inputs from all relevant stakeholders including researchers, top management universities, government agencies, libraries, research funder organizations, publishers, legal units, industries and research managers.

1.2. THE RATIONALE FOR THE GUIDELINES

There remain some concerns and challenges in Open Science, and the way they are being accessed needs to be addressed. Among the issues raised include the potential misuse of the data, which could result in misunderstanding of the meaning and compromised quality of shared data either among researchers or the public. Other opinions highlighted the need to formulate a strong and robust National Policy and Guidelines on Open Science in Public Funded Research which unequivocally clarifies research data ownership, recognises the role of data stewards and data curators, provide rewards and incentives to data contributors as well as establish rigorous security and privacy standards for data sharing practices.

More importantly, for MOSP to be successful, each initiative must be organised holistically, integrated and coherent with the overall goal, understood across all levels especially by researchers, and involves everyone's participation. This is important to develop a strong coherence with the mission, the vision and strategic thrusts outlined in the National Policy on Science, Technology and Innovation (NPSTI 2021-2030). The Ministry of Science, Technology and Innovation (MOSTI) is responsible for the promulgation and implementation of the Guidelines.

1.3. PURPOSE

The Guidelines provide the best practices for applying Open Science and achieving its fundamental goals. The main objectives of the Guidelines are to:

a) Provide better management of research data;

b) Assist researchers in Malaysia to deposit and retain research data files and datasets, publications and records and to contribute to scientific advancements through its availability for sharing; and

c) Ensure that research data generated in the conduct of research activities in all institutions are managed in a systematic and comprehensive manner to ensure quality, integrity, accountability, long term availability, appropriate sharing and compliance with the requirements of funding agencies.

1.4. DEFINITION OF OPEN SCIENCE AND OPEN ACCESS

For the purpose of the Guidelines, the OECD's definition will be adopted as follows:

Open Science is defined as "efforts by researchers, governments, research funding agencies or the scientific community itself to make the primary outputs of publicly funded research results – publications and the research data – publicly accessible in digital format with no or minimal restriction as means for accelerating research; these efforts are in the interest of enhancing transparency and collaboration and fostering innovation.".

Open Access is defined as the right to read, download and print as well as the right to copy, distribute, search, link, crawl and mine. However, the accessibility based on the rights must be subscribed and be subjected to the FAIR principle.

1.5. ADOPTION OF FAIR PRINCIPLES

Open science should be principle-based but be adapted to local realities. Based on the above definitions, the principles of **Findable, Accessible, Interoperable and Reusable (FAIR)** will be adopted under this Guidelines. FAIR data means that data is not always open, but it should be as open as possible, and as closed as necessary. With FAIR data, researchers can create, share and reuse quality, valuable, high integrity and responsible data, fuelling scientific progress to its fullest potential.

---

2 There are other main acceptable definitions of Open Science globally and are as follows:

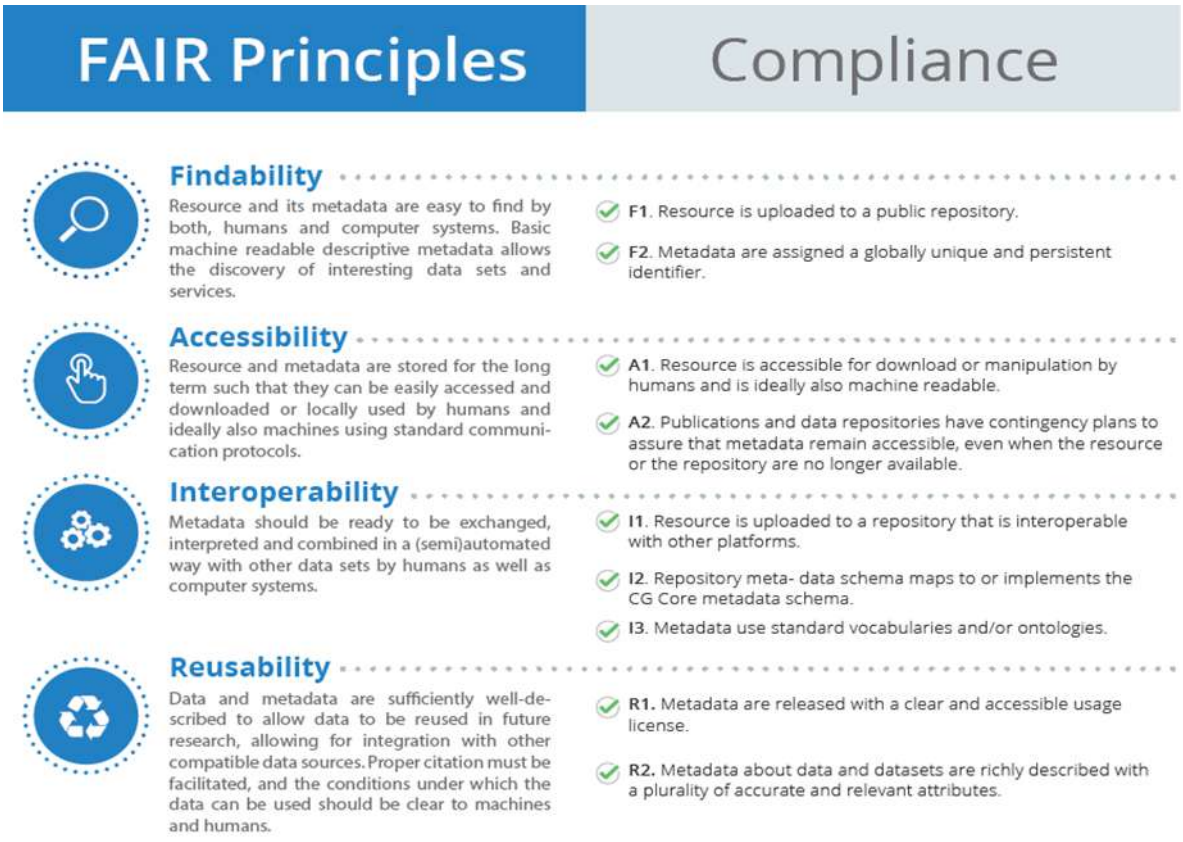
a) OECD2: "...efforts by researchers, governments, research funding agencies or the scientific community itself to make the primary outputs of publicly funded research results – publications and the research data – publicly accessible in digital format with no or minimal restriction as means for accelerating research; these efforts are in the interest of enhancing transparency and collaboration, and fostering innovation.".

b) FOSTER3: "...the practice of science in such a way that others can collaborate and contribute, where research data, lab notes and other research processes are freely available, under terms that enable reuse, redistribution and reproduction of the research and its underlying data and methods.".

c) RAND Corporation4: "Open Science refers to ongoing changes in the way research is conducted: for scientists themselves, through increasing the use of open access scientific publishing and open data, and for the public, through increasing their understanding of and participation in science ... Open Science is one of three priority areas for European research, science and innovation policy.".

- **Findable** means that data and metadata are easily found by both humans and computers. Usually, this task is enabled by machine-readable persistent identifiers and metadata.
- **Accessible** means that data can be retrieved using the outline protocols by appropriate people, at an appropriate time and in an appropriate way. Data can be FAIR even if the data has various levels of accessibility, such as: (1) Data is completely private; (2) Data is accessible by a defined group of people; and (3) Data is accessible by everyone.
- **Interoperable** means that the terminology system, protocols, standards and formats built and employed for datasets that are stored in a platform can be used and can communicate with other tools or platforms.
- **Reusable** means that data is well-defined and can be used for different purposes and in different settings, and the legal use is regulated by different terms and conditions. Data can be FAIR even if the data has various levels of reusability, depending on the stipulated licensing terms (E.g. acknowledgement, access and methods of data re-use, charges, exemption use of personal, sensitive and restricted data and proprietary information).

BOX 1.1. FAIR PRINCIPLES



Source: Horizon 2020, European Commission

1.6. ABOUT THE GUIDELINES

a) The Guidelines on Open Science in Public Funded Research sets out the directions in spurring Open Science movement in Malaysia. This document serves as a national guideline that defines types of published or documented research data including raw data that can be shared under specified conditions, harmonises definitions and terminologies, and outlines incentives for data sharing to promote the culture of openness where raw research data are shared among research from various disciplines in Malaysia, and to support good practice for raw research data sharing.

b) The Guidelines also develops guidelines for the management of open data sharing and research data management plan that outlines how research data arising from the research project will be handled during and after the project is completed, by describing what data will be shared and/or made open, and how it will be curated and preserved to ensure that the raw research data is accessible beyond the life of the project. The Malaysia Open Science Platform (MOSP) encourages all researchers to prepare a data management plan for publicly funded research projects to ensure that the raw research data generated by research projects are deposited at institutional repositories or data publishing partners' repositories and all metadata are permanently archived in the Malaysia Open Science Platform.

1.7. APPLICATION OF GUIDELINES

This Guidelines shall apply to:

(a) all staff, researchers, students and any other persons involved in the design, conduct, administration, or reporting of research performed at or under the auspices of Malaysian Universities, Research Institutes, and the Government Entities including consultants and visiting researchers.

(b) all research activities conducted in all universities, research institutes and government entities that received funding from the Malaysian government.

(c) all stages of the data life cycle- before, during and after. The Guidelines will operate in conjunction with other related national and institutional policies and guidelines and the Malaysian Laws and government policies.

The Guidelines shall not apply to:

(a) Research activities conducted for third party organisations using private or international funding.

(b) Consultancy services conducted for third party organisations including work carried out using or by Malaysian Universities, Research Institutes and the Government Entities' research facilities.

The National Guidelines on Open Science is prepared by referring to the existing Policy and Guidelines for Open Science which were developed in other countries, as well as based on the relevant laws, policies and regulations related to Open Science in Malaysia. Since the Guidelines is intended to first raise awareness and establish early adopters of Open Science, it is so designed to ensure the realisation of Open Science in the country at this stage. The Guidelines will be revised as and when necessary. In making Open Science though Open Access a reality, the current Guidelines will adopt three kinds of measure: mandatory rules (sticks), incentive mechanisms (carrots), and "enablers (soft and hard infrastructure)" as follows:

a) **Mandatory rules:** compelling open data sharing in Open Science a compulsory requirement or a prerequisite in research grant agreements or spelt out in national strategies or institutional policy.

b) **Incentive mechanisms:** incentivising those that are involved or celebrating or promoting Open Data Sharing in Open Science in the form of a financial incentive such as to cover the cost related to "open access publishing or the release of data sets.". The incentive can also come in non-financial form such as awarding recognition or official acknowledgment (or even involving career advancement) to researchers and academicians who are strong advocates for open science.

c) **Enablers:** involves the development of soft and hard infrastructure such as building a trusted sharing platform on Open Science (that allows ease of sharing of scientific articles or research data), sets of skills and training etc.

Since Open Science is defined as a global movement to make **scientific research, data and dissemination accessible** to all levels of an inquiring society, amateur or professional, the focus of the Guidelines on Open Science will be on open access to scientific research documents and research data including raw data in public funded research.

1.8. DATA SHARING FRAMEWORK IN OPEN SCIENCE

The framework is based on the 3 measures and 2 main areas as depicted in **Figure 1.1**. It will describe the **rewards and incentives** to research and data contributors as well as the best practices they needed to adopt together with the established rigorous security and privacy standards they need to conform or adhere to as part of the **mandatory rules** since open access does not mean "open to all" or that everything needs to be opened or disclosed.

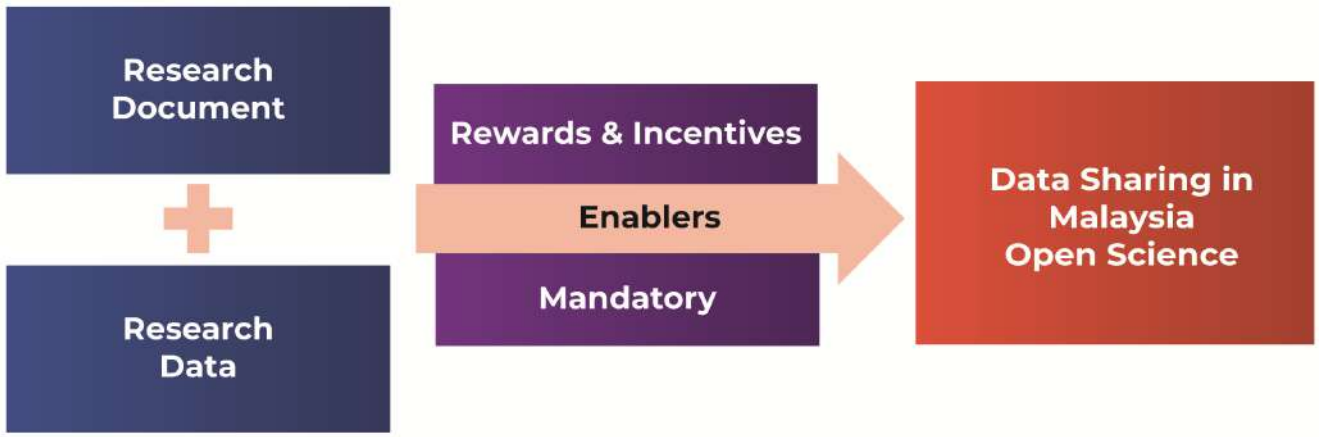


Figure 1.1. Data Sharing Framework in Open Science

CHAPTER 2

OPEN ACCESS TO SCHOLARLY PUBLICATIONS AND RESEARCH DATA

2.1. BACKGROUND

The Guidelines explain the rules on open data sharing through open access to scientific research documents and research data that beneficiaries have to follow in projects funded under government or public funds. **Figure 2.1**. describes the general routes practised and adopted.[3]. The Guidelines will adopt a similar route in the open data sharing process.

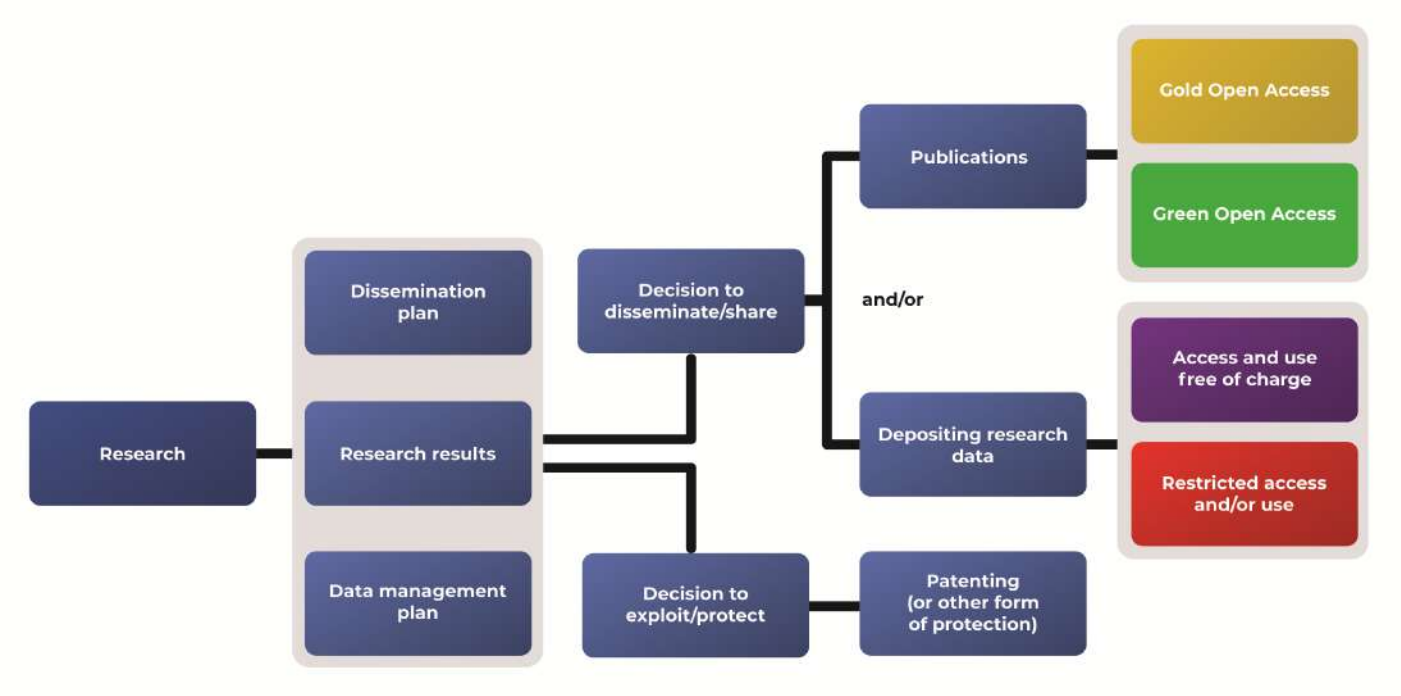


Figure 2.1. Flow process of Open Access to Scholarly Publications and Research Data
Source: Horizon Online Manual 2020, European Commission

2.2. PRINCIPLES

Though Open Science is based on 4 broad principles of FAIR, the Guidelines spelt out here are based on a set of specific principles adopted and adapted for this purpose. These principles are.[4]:

[3] by many organisations in particular the European Commission on open access to scientific research publications and research data.

[4] The principles and Guidelines are based on OECD document: "OECD Principles and Guidelines for Access to Research Data from Public Funding.". In January 2004, 30 OECD countries including China, Israel, Russia and South Africa agreed to adopt a "Declaration on Access to Research Data from Public Funding.". In recognising the significance of access to research data, OECD was asked "to develop a set of OECD guidelines based on commonly agreed principles to facilitate optimal cost-effective access to digital research data from public funding to be endorsed by the OECD Council at a later stage". In October 2006 the OECD's Committee for Scientific and Technological Policy approved the principles and guidelines and was endorsed by the OECD Council on 14th December 2006.

A. **Openness** – implies access should be on equal terms, easy, timely, user-friendly and preferably Internet-based.

B. **Flexibility** - taking into account the VUCA (vulnerability, uncertainty, complexity and ambiguity) local, regional and global events or situations on Open Science and is tailor-made based on the mutual "needs and offerings" of parties involved.

C. **Transparency** – means availability of data in a transparent way, preferably via the Internet. It includes research data that are easily findable and accessible via the Internet besides public research entities who are actively disseminating information on research data to individual researchers, academic associations, universities and other stakeholders.

D. **Legal and moral conformity** - Data access should conform to legal and moral rights of all stakeholders with some restriction to access of research data on the basis of national security (data intelligence, military activities etc), privacy and confidentiality (data on human subjects and other personal data that are subject to national privacy laws), trade secrets and intellectual property rights, protection of rare, threatened or endangered species for protection and conservation and legal data under legal actions.

E. **Protection of intellectual property** - Data access should not violate copyright or of other intellectual property laws relevant to publicly funded research databases.

F. **Formal responsibility** – implies that access to data should be formalised in terms of institutional practices on data-related activities such as "authorship, producer credits, ownership, dissemination, usage restrictions, financial arrangements, ethical rules, licensing terms, liability, and sustainable archiving.".

G. **Professionalism** relates to management of research data based on the professional standards and values embodied in the codes of conduct of the scientific communities involved such as the use of codes of conduct for professional scientists and their communities, mutual trust between relevant parties (researchers, institutions and other stakeholders).

H. **Interoperability** – one of the main principles of FAIR which relates to "technological and semantic interoperability" and is considered as a key consideration in enabling and promoting international and interdisciplinary access to and use of research data. Interoperability means that standards must be clearly described.

I. **Quality** involves "value and utility of research data" which is highly dependent on the quality of the data itself. Ensuring compliance to quality standards should be adhered to by data managers, and data collection organisations, if available, since "universal data quality standards" are occasionally not useful or realistic. However, measures should be taken to ensure good practices are adhered to (such as methods, techniques and instruments employed in data collection, dissemination and accessible) in "safeguarding quality and authenticity" in terms of origin of sources.

J. **Security** implies the guarantee of integrity and security of research data. Factors like completeness of data and absence of errors affect integrity. As for security, data protection through "intentional or unintentional loss, destruction, modification and unauthorised access in conformity with explicit security protocols" should be given strategic priority apart from the safeguarding of data storage sets and equipment "from environmental hazards such as heat, dust, electrical surges, magnetism, and electrostatic discharges.".

K. **Efficiency** is one key goal in promoting data access and sharing "to improve the overall efficiency of publicly funded scientific research to avoid the expensive and unnecessary duplication of data collection efforts.". Some considerations are given to ensure its cost effectiveness in retaining data through cost-benefit assessments conducted periodically to ensure that the data sets with the greatest potential utility are preserved and made accessible.

L. **Accountability** implies that data access is subjected to "periodic evaluation by user groups, responsible institutions and research funding agencies" since such evaluations will assist to step up the support of open access among the relevant stakeholders.

M. **Sustainability** relates to long term retention to access of publicly funded research data. Given that most research projects, and the public funding provided are limited in terms of duration, it can be very challenging to ensure long term access to the data produced. The best the research funding agencies and research institutions could do in long term preservation of data is, at the outset of each new project, to determine the most appropriate archival facilities for the data.

2.3. DATA SHARING THROUGH ACCESS OF SCHOLARLY PUBLICATION

What does Open Access to scholarly publications mean? Open Access to pre-processed and processed research results allows materials to be found or obtained via an internet search and be made available free of charge and free for further reuse. In providing accessibility, however, there are different routes of providing access to scientific articles, and they can be categorised as gold, green, diamond, and hybrid and bronze publishing. As for the term 'gold', it can be applied in different ways,[5], though in this Guidelines it is used as defined below:

> "1) the researcher publishes an article in an open access journal (gold);
>
> 2) the researcher publishes in a traditional subscription-based publication and thereafter a copy of the manuscript is published in open access via a digital archive as soon as the publication permits this (green/self-archiving).
>
> 3) the researcher publishes the article in a traditional subscription-based publication and, for a fee, the article is made open access with immediate effect (hybrid). Researchers who publish their results in book form or in the form of an artistic work can also make available the material on an open access basis [6].

2.3.1. Routes to Open Access to Scholarly Publications

There are 2 main routes to open access. They are:

a. **Self-archiving /'green' open access** – a version of an author's manuscript into a repository, making it freely accessible for everyone. The version can be deposited into a repository and is dependent on the funder or publisher.

b. **Open access publishing/'gold' open access** - an author publishes an article in an on-line open access journal. In this model, the payment of the publication costs is shifted away from the subscribing readers. The most common business model is based on on-off payments by authors. These costs, often referred to as Article Processing Charges (APCs) are usually borne by the researcher's university or research institute or the agency funding the research.

In other cases, the costs of open access publishing are covered by subsidies or other funding models.

2.3.2. Open Access to Scholarly Publication and Scientific Records

> 1. The Guidelines requires that a machine-readable electronic copy of the published version or final peer-reviewed manuscript accepted for publication of all peer reviewed publications produced as a result of research supported, either in entirety or in part by a Research Funder Organisation, is deposited in a suitable Open Access repository. Deposit should be made immediately upon acceptance for publication and the metadata made fully open, searchable and machine-readable from the time of deposit. This step also applies in the case of Open Access publishing ("Gold Open Access").

---

[5] The EU Commission defines gold as "immediate open access that is provided by a publisher" a formulation that does not exclude hybrid publishing. http://ec.europa.eu/digital-agenda/en/open-access-scientific-information 27/11/2014.

2. In the case of "Green Open Access", the Guidelines requires that the full-text of all such publications be made available under a standard open licence immediately where possible and in any case no later than 6-months after publication in Science, Technology, Engineering and Mathematics (STEM) or 12 months after publication in the Social Sciences and Humanities (SSH). If a journal's permitted embargo period is longer than these, authors should either negotiate with the publisher to retain the rights so as to comply with this policy, or find a journal that enables them to comply without the need for negotiation. The Guidelines encourages retaining ownership of copyright and to licence to publishers only those rights necessary for publication. This is possible through the use of addenda to the publishing contract.

3. The Guidelines will recognize compliant journals as those that adhere to the above provisions. The hybrid model of publishing is not compliant with the above principles and related costs will not be considered as eligible.

4. The Guidelines will recognise Open Access publication fees such as Article Processing Charges (APCs) or Book Processing Charges (BPCs) as eligible costs according to the funding guidelines. For quality assurance purposes, eligible journal titles must be listed in standard directories like the Directory of Open Access Journals (DOAJ) or PubMed.

5. The Guidelines requires that funded publications must be made available under an open content licence, such as Creative Commons (CC BY). In all cases, the licence applied should fulfil the requirements defined by the Berlin Declaration.

6. While the dominant type of scientific publication is the journal article, grantees are strongly encouraged to provide Open Access to other types of publications such as monographs, book chapters, conference proceedings, grey literature, reports, etc.

2.4. OPEN ACCESS TO RESEARCH DATA

Open Science is commonly misunderstood in the sense that all data has to be open by default. However, there is a wide spectrum of data sharing practices from entirely making the data open, to restricted and closed access. In the cases of government funded projects, raw research data and datasets that belong to public domains and are useful for collaborative solutions, such as COVID-19, must be made open and available, while also observing legal provisions for data sharing. For restricted or closed raw research datasets, the metadata should be published to indicate that a particular research has been done and the raw research datasets have been gathered and consolidated.

2.4.1. Research Data Lifecycle

The research data lifecycle and the six stages (data acquisition, processing, analysis, curation, sharing and data reuse) involved in it is depicted by **Diagram 2.2.**

The data acquisition stage is where all the sensors, data streams, data repositories are connected for the use of researchers. Data processing involves computing platform or workbench for users to use or catalogue features that allow the researchers to choose a particular dataset and add to their data cut and send for data processing [8] . In the data analysis stage, the data scientist or analyst will draw a conclusion from the dataset. At all the three stages aforementioned, the project or the data is still active so the datasets will keep changing. The data must undergo a curation process before it can be shared to others.

---

[7] Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities (2003) https://openaccess.mpg.de/Berlin-Declaration.

[8] data cut feature in Edinburgh and whether to implement this data cut feature into MOSP is still questionable

Data stewards play a crucial role in guiding the data originators in managing their data to assure the data is in good quality and is preserved. From this stage onwards, the data is all set to be shared and reused by others. Here is where the data repositories or domain repositories whether owned by individual institutions or shared repositories by an independent party come into place. For example, NCBI has its gene bank where all the gene datasets are uploaded before paper publication. Each data repository allows the users to search the database, for instance government data can be accessed in 'data.gov.my' and all institutions data under MOSTI can be accessed in 'radars.mosti.gov.my'. However, the challenge appears when a user wants to search for a dataset across all these repositories under the same gateway. Hence, the moderator pointed out if we should have a registry to harvest the data from data repositories and put them into one place (data catalogue) and to be maintained, for instance by MOSP. The discovery service will then allow users to search for any data across all these repositories. At the final stage of the research data lifecycle, the data should be allowed to be reused.

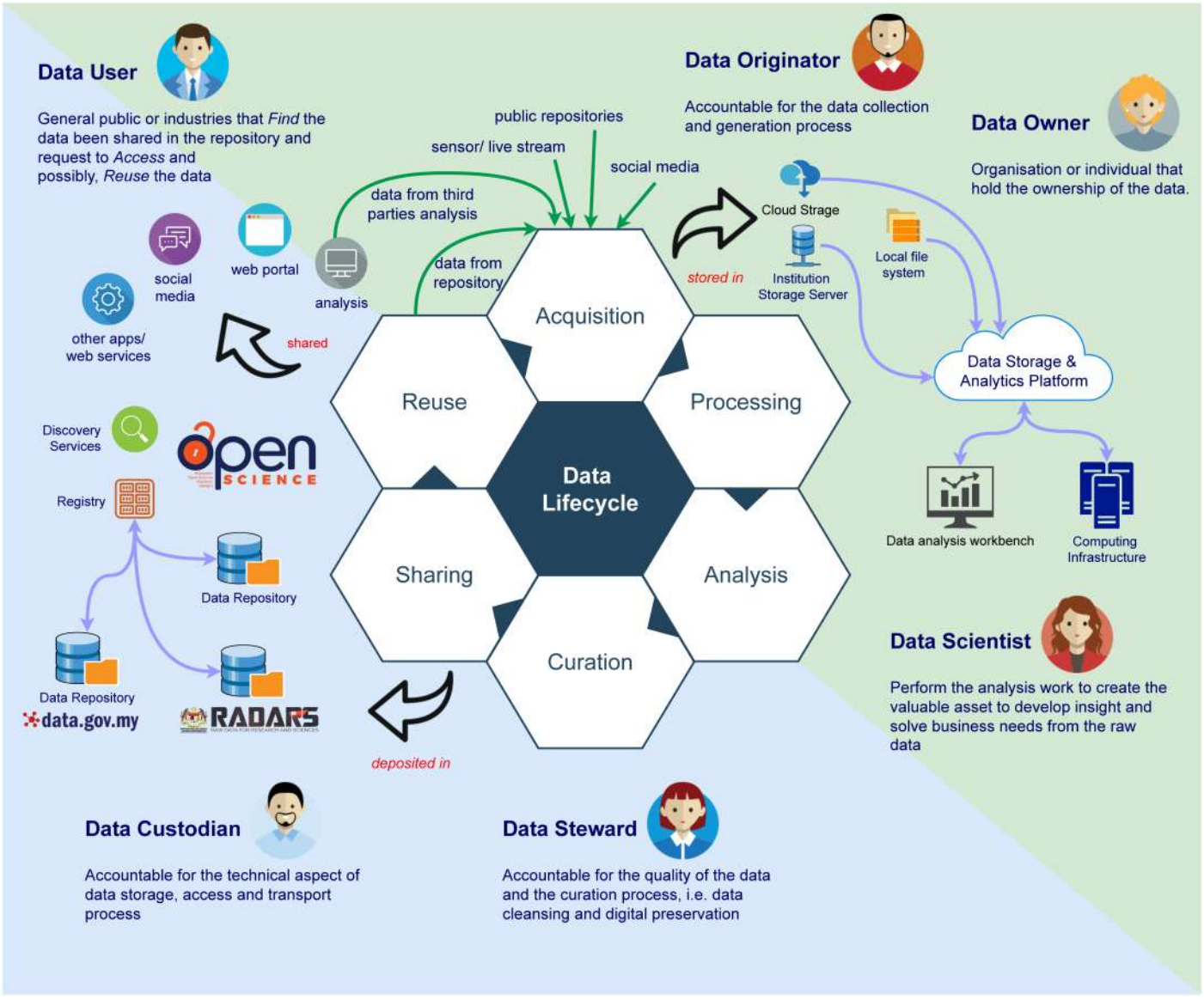


Figure 2.2. Research Data Lifecycle Source: MOSP, 2020

2.4.2. Guidelines to Open Access to Research Data

1. Requires researchers to deposit the research data (which includes raw data) that were used and processed to yield the results that are published in scientific publications in institutional repositories. Research data should be assigned with persistent identifiers.

2. Requires that research data and services are handled according to FAIR principles (i.e. Findable, Accessible, Interoperable and Re-usable). Raw research data should also be traceable and whenever possible, available for subsequent use.

3. The institutional repository follows the principle “as open as possible as closed as necessary”. If data cannot be open due to legal, privacy or other concerns, this should be clearly explained. Metadata ensuring that data are findable should be provided in all instances.

4. Encourages the adoption of the MOSP requirements for monitoring of Open Science resources.

5. Requires researchers to submit a DMP to the appropriate service for every research activity they are involved in.

6. Requires researchers to define post-project usage rights through the assignment of appropriate licences

7. Requires that data are stored for a period as defined by the respective communities.

8. The minimum archive duration for research data is 10 years after the assignment of a persistent identifier. In the event that these records need to be deleted or destroyed after the expiration of the required archived duration or for legal and ethical reasons, such actions need to consider all legal and ethical perspectives.

9. All costs associated with the management of research data are considered eligible costs under the Guidelines. However, data management costs should be indicated or specified in the grant agreement application together with the data management plan created by the applicants.